

Deterministic MST Sparsification in the Congested Clique

Janne H. Korhonen

School of Computer Science, Reykjavík University

Abstract. We give a simple deterministic constant-round algorithm in the congested clique model for reducing the number of edges in a graph to $n^{1+\varepsilon}$ while preserving the minimum spanning forest, where $\varepsilon > 0$ is any constant. This implies that in the congested clique model, it is sufficient to improve MST and other connectivity algorithms on graphs with slightly superlinear number of edges to obtain a general improvement. As a byproduct, we also obtain a simple alternative proof showing that MST can be computed deterministically in $O(\log \log n)$ rounds.

1 Introduction

MST in the congested clique. The *congested clique* [5] is a specialisation of the standard CONGEST model of distributed computing; in the congested clique, each of the n nodes of the network can send a different message of $O(\log n)$ bits to each other node each synchronous communication round. The congested clique model has attracted considerable interest recently, as the fully connected communication topology allows for much faster algorithms than the general CONGEST model.

Minimum spanning tree is perhaps the most studied problem in the congested clique model, and a good example of the power of the model. The Lotker et al. [5] paper introducing the congested clique model gave an $O(\log \log n)$ -round deterministic MST algorithm. Subsequently, even faster randomised algorithms have been discovered: the $O(\log \log \log n)$ -round algorithm by Hegeman et al. [2], and the recent $O(\log^* n)$ -round algorithm by Ghaffari and Parter [1].

MST sparsification. Both of the above fast randomised MST algorithms are based on fast randomised *graph connectivity* algorithms. To solve MST, they use a reduction of Hegeman et al. [2] from MST to graph connectivity; this works by (1) reducing general MST into two instances of MST on graphs with $O(n^{3/2})$ edges using a randomised sampling technique of Karger et al. [3], and (2) reducing MST on sparse graphs to multiple independent instances of graph connectivity.

In this work, we take a closer look at the sparsification step of the Hegeman et al. [2] reduction. Specifically, we show that it is possible to do obtain much stronger sparsification for connectivity problems in constant rounds without using randomness:

Theorem 1. *Given a weighted graph $G = (V, E)$ and an integer k , we can compute in $O(k)$ rounds an edge subset $E' \subseteq E$ with $|E'| = O(n^{1+1/2^k})$ such that E' contains one minimum spanning forest of G .*

In particular, Theorem 1 implies that graphs with slightly superlinear number of edges are the hardest case for connectivity problems in the congested clique, as graphs with linear number of edges can be learned by all nodes in constant rounds using the routing protocol of Lenzen [4]. Alas, our sparsification technique alone fails to improve upon the state-of-the-art even for deterministic MST algorithms, though applying Theorem 1 with $k = \log \log n$ does give an alternative deterministic $O(\log \log n)$ algorithm for MST in the congested clique.

2 Deterministic MST Sparsification

Let $\mathcal{S} = \{S_1, S_2, \dots, S_\ell\}$ be a partition of V . For integers i, j with $1 \leq i \leq j \leq \ell$, we define

$$E_{ij}^{\mathcal{S}} = \{\{u, v\} \in E : u \in S_i \text{ and } v \in S_j\},$$

and denote by $G_{ij}^{\mathcal{S}}$ the subgraph of G with vertex set $S_i \cup S_j$ and edge set $E_{ij}^{\mathcal{S}}$.

Definition 2. For $0 < \varepsilon \leq 1$, graph $G = (V, E)$ and partition $\mathcal{S} = \{S_1, S_2, \dots, S_\ell\}$ of V , we say that (G, \mathcal{S}) is ε -sparse if $\ell = n^\varepsilon$, each $S \in \mathcal{S}$ has size at most $n^{1-\varepsilon}$ and for each i, j with $1 \leq i \leq j \leq \ell$, we have $|E_{ij}^{\mathcal{S}}| \leq 2n^{1-\varepsilon}$.

If (G, \mathcal{S}) is ε -sparse, then G can have at most $2n^{1+\varepsilon}$ edges. Moreover, we will now show that we can *amplify* this notion of sparseness from ε to $\varepsilon/2$ in constant rounds. Observing that for any graph $G = (V, E)$ and $\mathcal{S} = \{\{v\} : v \in V\}$, we have that (G, \mathcal{S}) is 1-sparse, we can start from arbitrary graph and apply this sparsification k times to obtain sparsity $1/2^k$, yielding Theorem 1.

For convenience, let us assume that the all edge weights in the input graph are distinct, which also implies that the minimum spanning forest is unique. If this is not the case, we can break ties arbitrarily to obtain total ordering of weights. Recall that each node in V receives its incident edges in G as input.

Lemma 3. *Given a graph $G = (V, E)$ with distinct edge weights and unique MSF $F \subseteq E$, and a globally known partition \mathcal{S} such that (G, \mathcal{S}) is ε -sparse, we can compute a subgraph $G' = (V, E')$ of G and a globally known partition \mathcal{T} such that (G', \mathcal{T}) is $\varepsilon/2$ -sparse and $F \subseteq E'$ in constant number of rounds.*

Proof. To obtain the partition $\mathcal{T} = \{T_1, T_2, \dots, T_{n^{\varepsilon/2}}\}$, we construct each set T_i by taking the union of $n^{\varepsilon/2}$ sets S_j . Clearly sets T_i constructed this way have size $n^{1-\varepsilon}$, and this partition can be constructed by the nodes locally. Since (G, \mathcal{S}) is ε -sparse, we now have that

$$|E_{ij}^{\mathcal{T}}| = \sum_{x: S_x \subseteq T_i} \sum_{y: S_y \subseteq T_j} |E_{xy}^{\mathcal{S}}| \leq (n^{\varepsilon/2})^2 2n^{1-\varepsilon} = 2n.$$

We assign arbitrarily each pair (i, j) with $1 \leq i \leq j \leq n^{\varepsilon/2}$ as a *label* for distinct node $v \in V$. The number of such pairs (i, j) is at most $(n^{\varepsilon/2})^2 = n^{\varepsilon} \leq n$, so this is always possible, though some nodes may be left without labels. The algorithm now proceeds as follows:

1. Distribute information about the edges so that node with label (i, j) knows the full edge set $E_{ij}^{\mathcal{T}}$. Since $|E_{ij}^{\mathcal{T}}| \leq 2n$, this can be done in constant rounds using the routing protocol of Lenzen [4].
2. Each node with label (i, j) locally computes a minimum spanning forest $F_{ij}^{\mathcal{T}}$ for the subgraph $G_{ij}^{\mathcal{T}}$ using information obtained in previous step. Since $|T_i \cup T_j| \leq 2n^{1-\varepsilon/2}$, we also have $|F_{ij}^{\mathcal{T}}| \leq 2n^{1-\varepsilon/2}$.
3. Redistribute information about the sets $F_{ij}^{\mathcal{T}}$ so that each node knows which of its incident edges are in one of the sets $F_{ij}^{\mathcal{T}}$. Again, this takes constant rounds.

Taking $E' = \bigcup_{(i,j)} F_{ij}^{\mathcal{T}}$, we have that (G', \mathcal{T}) is $\varepsilon/2$ -sparse. To see that E' also contains all edges of F , recall the fact that an edge $e \in E$ is in MSF F if and only if it is the minimum-weight edge crossing some cut (V_1, V_2) , assuming distinct edge weights (see e.g. Karger et al. [3]). If edge $e \in E_{ij}^{\mathcal{T}}$ is in F , then it is minimum-weight edge crossing a cut (V_1, V_2) in G , and thus also minimum-weight edge crossing the corresponding cut in $G_{ij}^{\mathcal{T}}$, implying $e \in F_{ij}^{\mathcal{T}}$. \square

Acknowledgements. We thank Magnús M. Halldórsson, Juho Hirvonen, Tuomo Lempinen, Christopher Purcell, Joel Rybicki and Jukka Suomela for discussions, and Mohsen Ghaffari for sharing a preprint of [1]. This work was supported by grant 152679-051 from the Icelandic Research Fund.

References

- [1] Mohsen Ghaffari and Merav Parter. MST in log-star rounds of congested clique. In *Proc. 35nd ACM Symposium on Principles of Distributed Computing (PODC 2016)*, 2016.
- [2] James W. Hegeman, Gopal Pandurangan, Sriram V. Pemmaraju, Vivek B. Sardeshmukh, and Michele Scquizzato. Toward optimal bounds in the congested clique: Graph connectivity and MST. In *Proc. 34nd ACM Symposium on Principles of Distributed Computing (PODC 2015)*, pages 91–100, 2015. doi:[10.1145/2767386.2767434](https://doi.org/10.1145/2767386.2767434).
- [3] David R. Karger, Philip N. Klein, and Robert E. Tarjan. A randomized linear-time algorithm to find minimum spanning trees. *Journal of the ACM*, 42(2):321–328, 1995. doi:[10.1145/201019.201022](https://doi.org/10.1145/201019.201022).
- [4] Christoph Lenzen. Optimal deterministic routing and sorting on the congested clique. In *Proc. 32nd ACM Symposium on Principles of Distributed Computing (PODC 2013)*, pages 42–50, 2013. doi:[10.1145/2484239.2501983](https://doi.org/10.1145/2484239.2501983).
- [5] Zvi Lotker, Boaz Patt-Shamir, Elan Pavlov, and David Peleg. Minimum-weight spanning tree construction in $O(\log \log n)$ communication rounds. *SIAM Journal on Computing*, 35(1):120–131, 2005. doi:[10.1137/S0097539704441848](https://doi.org/10.1137/S0097539704441848).